DOCUMENT RESUME

ED 448 747 IR 020 510

AUTHOR Bogen, Manfred; Bonkowski, Christian; Borowski, Marion; Loffler, Jobst

TITLE Digitizing a Cultural Heritage--The Key Issue for Preservation and Electronic Publishing.

PUB DATE 2000-11-00

NOTE 7p.; In: WebNet 2000 World Conference on the WWW and Internet Proceedings (San Antonio, TX, October 30-November 4th, 2000); see IR 020 507.

PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS Electronic Libraries; Electronic Publishing; Evaluation Criteria; *Evaluation Methods; Foreign Countries; *Manuscripts; *Museums; Musical Composition; Musicians; Optical Scanners; Pilot Projects; *Preservation; Vendors

IDENTIFIERS Beethoven (Ludwig van); *Digitizing; Germany; Music Composers

ABSTRACT

This paper describes a realization approach to evaluate the best fitting service provider in the context of "The Digital Beethoven House" pilot project, involving the digitizing of a composer-referred collection of colored manuscripts and other sources at the Beethoven House Association (Germany). Practical hints for museums are provided on how to find the right scan service providers, how to do the digitizing of their museum's stock right, and how to receive useful output for the museum's Internet offer. The first section presents a methodology for finding the right scan service provider, including the following steps: (1) development of quality criteria for the whole digitization chain; (2) market analysis of scanning services and call for tender; (3) evaluation of the offers; (4) visits of the best qualified vendors; and (5) decision making. The second section describes quality criteria for digitizing paper related to the quality of scans, digital master, scan equipment, and experiences of the scanning service provider. The third section summarizes the most common publishing scenarios for managing digital documents, including digital library and Internet publishing, CD-ROM production, facsimiles and posters, and virtual reality. The fourth section highlights digitizing other media, including audio, video, and 3D objects. (Contains 19 references.) (MES)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 448 747

Digitizing a Cultural Heritage - The Key Issue for Preservation and Electronic Publishing

Manfred Bogen, Christian Bonkowski, Marion Borowski, Jobst Löffler
GMD – German National Research Center for Information Technology,
Schloss Birlinghoven, D-53754 Sankt Augustin
E-mail: {Manfred.Bogen | Christian.Bonkowski | Marion.Borowski | Jobst.Loeffler}@gmd.de

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

G.H. Marks

Abstract: Digitizing is the first and most important step in a preservation and electronic publishing process applied to analog information available today as text, graphics, or multimedia. In most cases, it cannot be done by museums alone.

In this paper we describe our realization approach to evaluate the best fitting service provider in the context of 'The Digital Beethoven House' project. It was a difficult task because the digitizing of a composer-referred collection of colored manuscripts and other sources was a pilot project for all participants.

The paper gives practical hints for museums how to find the right scan service providers, how to do the digitizing of their museum's stock right and how to receive useful output for the museum's Internet offer.

Introduction: 'The Digital Beethoven House' Project

Since 1889, the Beethoven House Association in Bonn maintains as private carrier at Beethoven's birthplace a museum for the life and works of Ludwig van Beethoven (1770-1827). Beside the museum and its collections, a Beethoven Archive as scientific research institute, a publishing house and the chamber music hall H. J. Abs belong to this unique ensemble.

Only a small part of the collection can be shown for preservation reasons or from lack of space in the permanent exhibition or in special exhibitions at the Beethoven House.

The project 'The Digital Beethoven House', in which GMD participates as technology partner, starts here. The major target of the project is it to enable a more intensive occupation with life and works of Ludwig van Beethoven for a broader interested audience [Bogen & Lahme & Lechner 99].

It is an innovative digital library project, in which for the first time worldwide a composer-referred collection is made accessible by multimedia. For the first time music will be visualized in a virtual installation, for the first time music handwriting and other composer-referred documents will be digitized in colors. Naturally, 'The Digital Beethoven House' will be Online in the Internet.

This paper is structured as follows: The next section introduces our methodology to find the right scan service provider for 'The Digital Beethoven House' project. It is followed by a chapter that describes our respective requirements in more detail. The further section is about managing digital documents for publication purposes. The paper is complemented by some remarks on digitizing audio, video, and 3D models. Finally, the last section concludes the paper with some general remarks on the results so far and those expected to be achieved until the WebNet'2000 conference.

Realization Approach

The largest cultural treasures are situated behind fireproof walls and steel doors, protected against light and humidity, air-conditioned and dust free. The Beethoven House Association in Bonn e.g., store over thousand inestimably valuable manuscripts of the composer Ludwig van Beethoven: music handwritings [Fig. 1], letters [Fig. 2], sketches of sheets of music, first prints, colored engravings [Fig. 3], photographs, pictures. 9,000 colored pages of them and more than 16,000 pages in gray tones, with formats that go from DIN A6 to DIN A2, in parts even bound, are going to be digitized. But several questions arise:

- How does a museum achieve a future-proof digital archive, respectively digital copies with highest image quality and similarity to the original, which are applicable to various future purposes?

IR020510

BEST COPY AVAILABLE

- Is the own personnel qualified to execute the scanning work, in order to be prepared for later extensions?
- Or which image service bureau will meet the requirements for professional scans in color?

In the case of the 'The Digital Beethoven House' it was decided to charge a professional image service provider: Reproduction specialists, who are characterized by their daily work with high-quality colored historical collections. GMD as technology partner and consultant in 'The Digital Beethoven House' project was charged to find a competent scan service. The following methodology proved successful in doing so.
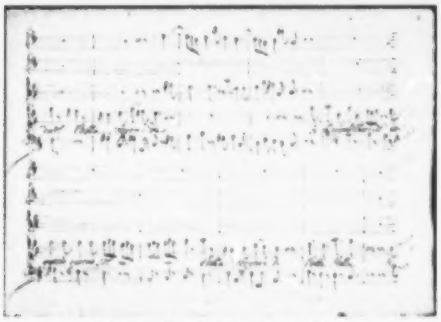
Figure 1: Music handwriting

Figure 2: A letter of Beethoven

Step 1: Development of Quality Criteria for the Whole Digitization Chain

A qualified scanning process needs to meet the requirements of the special documents of the museum. The criteria for the whole process of electronic publishing were collected.

Step 2: Market Analysis of Scanning Services and Call for Tender

The market was searched for potential candidates who were invited to offer their services according to the quality criteria of step 1.

Step 3: Evaluation of the offers

Each offer was looked through carefully. The most promising companies were chosen for a detailed view.

Step 4: Visits of the Best Qualified Vendors

The best scanning vendors had the chance to present their color consistent scanning chain and their professionals while scanning special exemplary originals of the museum. Digital copies on CD-ROM and proof prints were taken with us to compare the results of the different service providers.

Step 5: Decision Making

For each vendor we evaluated the fulfillment of our criteria.

The results were sobering: 6 out of 15 scanning services responded to our call for tender, out of which we chose 3 for a detailed evaluation. There were no scanning services with long-term experiences in scanning colored composer-referenced collections! They had a lot of experiences with two-tone or grayscale documents but for colored manuscripts they were all still learning. None could present color fidelity for all components of the digitization chain.

We decided for a candidate with experience in the whole digital library environment who presented a comparison of 8 different scanner types in contrast, color, sharpness, and depth of focus, tested with a representative test chart of the "Landesarchiv Baden-Württemberg" (Fig. 5).

Quality Criteria for Digitizing Paper

The requirements for the digitization process orient themselves at the nature, the condition and the value of the source material. The prime motto is to scan only once and to touch the original never again. As output of the scanning process the digital images must be of preservation quality, in order to use them instead of the historic collections themselves.

In the following paragraphs the most substantial quality criteria our outlined. They lean against national and international scanning guidelines experiences [Kenney & Chapman 96], [Mittler 98], which we evaluate in the light of our practical experiences.

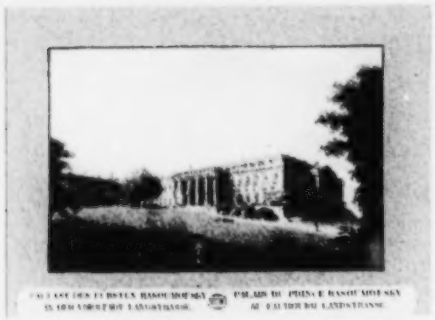

Figure 3: Colored engraving

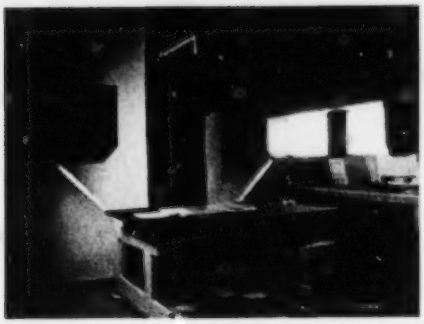

Figure 4: Highly qualified scan equipment

1. *Quality of Scans: Which image quality requirements are essential to achieve best quality copies?*

   Knowing that high quality extends the life of a digital collection, we decided for scans with at least 400-600 dpi and 48 bits color depth for colored material as well as for grayscale documents. We learnt that these are only corner points for the image quality, which cannot be forecasted with one or two features. *First, we know how many things can go wrong from the source material to the copy. Second, we know that in all cases we must give up something from the original [Chapman 99].* In the long run the practical demonstration and the digital outcome of a calibrated [ICC 93] digitization chain with sample charts must convince the musicologist of the similarity to the original. Likewise a compromise between image quality, filesize, and costs must be met.

2. *Digital Master: How largely, respectively how 'rich' is the digital master to become?*

   The digital master will not be transformed after scanning and captures all the information. As output format for this rich master we decided for the TIFF [TIFF 92] standard, which can be read from, written to and converted to by the common systems. The technical meta data will be stored within the header, in order to hold preservation information as close at the picture as possible, bibliographical and structural meta data will be stored as XML-files. As medium for the digital master and its ICC profiles [ICC 93] we chose CD-ROMs using an ISO standard [ISO9660 88] for storage which is not yet the case for DVDs. A storage concept will migrate these storage media to long-term solutions.

   Because of the administration overhead and the loss of quality for later digitizing, microfilms were no options for our work.

3. *Scan Equipment: Which scanner supplies the desired quality for all sizes and document types?*

   Several types of scanners are available for digitizing documents [Kenney & Chapman 96]. The appropriateness for a museum depends on the category of the archival material. *Physical handling is one of the most destructive things that can happen to a fragile object. One of the best ways to preserve it is to limit physical access to it [Noerr 98].* The originals have not to be damaged by contact, light or heat and must therefore be digitized as contact-free as possible. After the demonstration of different scan equipments and their results, digital cameras [Fig. 4] with book seesaws appeared to us best suiting for museums. They meet best the quality requirements for all document sizes and scan contact-free and face-up. The scanning speed is an essential and very cost consuming feature. Complete scans (incl. preparation) should not take more than 3-5 minutes.

4. *Experiences of a Scanning Service Provider: Which image service bureau is trustworthy?*

   The scan service is supposed to have reference projects where it handled similar archival documents which often are faded, torn, written on special paper and with ink, and even stitched. Reference projects should be verified by the customer. Experiences with grayscale or two-tone documents are not sufficient. The use of modern equipment by professionals taking into account the exclusiveness of the documents is decisive.

   Each step of the digitizing needs to be coordinated and documented by the scanning service provider to use the expensive equipment to capacity and to clearly define the interfaces to the customer.

Managing Digital Documents

The digital results must be of *functional* quality, so that the production of many types of derivatives is possible. The motto is to scan once and publish often. We present the most common publishing scenarios.

Digital Library and Internet publishing

After digitizing a museum is able to offer its treasures to the public via a Web based digital archive. Different access concepts will be realized, is it the minimal request for thumbnail images on screen or is it to offer full screen pictures for scientific workstations. In our application context three groups of users are going to be addressed. The Internet users have reading rights to the digital images with a lower resolution, but to the complete meta data. The Intranet users of the museum or the library have access to the material with higher quality as well as the visitors, travelling to Bonn and accessing the data via the public workstations in the Beethoven Salon, a real place on the premises of the Beethoven museum.

The digital master is the starting point for the subsequent treatment in order to serve the specific purposes. Its quality features described in the previous chapter make sure that each considerable size or format for current and future applications might be produced afterwards. The public never gets to face the large digital masters in the original format but rather its derivatives. On the basis of the ICC [ICC 93] profile the original RGB scan [Marshall], the *archival digital master*, is transformed to an image of the LAB color space [Marshall] with 24 bits color depth, the *processed digital master*. The processed digital master is converted to handy image formats, e.g. jpeg or forms of the wavelet [WAVE] compression which are more appropriate for the performant data transmission in Internet networks. It provides progressive loading, scalable sizes and quality that may be accessed with password protection. Other formats like flashpix [Rigby 96] are in discussion to meet the requirements of zooming of different image pieces.

The digital archive is part of our integrated digital library solution where the digital copies are brought together with already existing data providing captured information e.g. from the proprietary library systems.

CD-ROM Production

Digital copies of high quality (e.g. TIFF-format [TIFF 92]) will be used for Offline media like CD-ROMs which people may order to buy copies of their favorite documents. We chose the ISO 9660 [ISO9660 88] standard because of platform independence.

Facsimiles and Posters

The quality criteria for scanning are appropriate for the creation of high-qualified facsimiles reprints, the output of the digital master to the same analog medium as the original. Whether digital printing techniques with ICC profiles [ICC 93] are used or whether offset printing techniques are brought into action with manually adapted characteristic printing lines – musicologists should check the color fidelity and the whole appearance of the facsimiles on the used paper.

Enlargements for posters may be easily printed because a low resolution is sufficient for a distant view.

Virtual Reality

Some of the scanned material will be used to build 3D models for a virtual world, like Beethoven´s last study which visitors may experience in the context of the 'The Digital Beethoven House' [Bogen & Borowski 98]. For this virtual reality work scanned material with a resolution of 100 dpi is sufficient.

A Short Glimpse at Digitizing other Media

Collections of museums today contain more than just plain text documents and pictures. Multimedia material like audio, video or 3D objects have to be digitized and integrated too.

Audio: Most audio data are already digitally available on CDs. But these audio CDs cannot be used directly. They have to be converted to another digital format. For doing this a normal CD device in a computer together with special programs to read the audio tracks are enough.

However, proprietary audio data on analog media must be digitized differently. In spite of existing special hardware for this task, in most of the cases (e.g. small music pieces) a PC with a good soundcard and special programs are enough.

In 'The Digital Beethoven House' project only a few audio data from chamber music hall recordings have to be digitized. So a simple PC will be used for this purpose. The result is always a WAV file [RIFF]. This is the standard PC format but not a good solution for the WWW. Before converting the WAV files, it must be decided, whether streaming audio or download on the WWW should be provided. For streaming audio it is still the RealAudio format [REAL] to prefer. As a download format today MP3 [MP3 93] becomes more and more the standard.

Video: To digitize video data, there are many possibilities. It mainly depends on the source of the video data, how to digitize it, and on the future use of the digitized video.

In general, a video capture card is needed to digitize the analog format. As a standard output one or more Audio Video Interleaved (AVI [RIFF]) files are generated. The problem, which may occur, is that AVI files have a maximum size of 2 GB.

In a next step, the video can be edited, if needed. At the end, the AVI format has to be converted to other video formats. Which format is preferred, depends on the further use of the video. The most used formats are QuickTime [QUIC], MPEG [MPEG] and RealVideo [REAL]. All these formats have a loss of quality but a handy file size. The frame rate for the video compression can be selected. The standard for PAL is 25 frames per second (fps). This is mostly too much for the WWW.

To present the (streaming) video in the Web, a streaming server is needed. This solution is preferable to downloading the video.

3D Objects: Virtual 3D models of artwork, industrial products or any other physical object can be produced today with commercial equipment for 3D scanning with high accuracy. The objects are scanned by non-contact digitization methods; only light touches the objects. As an example the digitization of works of arts in a museum will be discussed here. The advantage for a museum, which possesses also digital models of its stock, is on the one hand the local and WWW based accessibility by art historians, educators and the public without strain for the artwork. On the other hand it is an important advantage to improve the accessibility by turning the viewing of art into an active rather than a passive experience. In addition to education, science and entertainment, the produced virtual models can be used for commercial and marketing purposes which is especially interesting for digital museums accessible through the WWW.

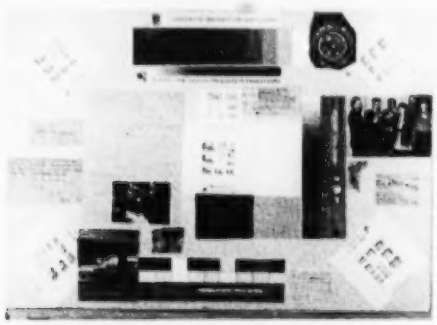

Figure 5: Test chart (Landesarchiv Baden-Württemberg)

Figure 6: 3D scanning of a Beethoven marble bust

The principle steps for the production of a virtual model and for making it accessible are 3D scanning, modeling with help of the resulting range image data, and visualization of the final virtual model. A laser triangulation scanner and a motorized gantry are used to produce a 3D polygonal surface mesh for objects of small dimensions as well as for very large objects like, for instance, the giant figure of Michelangelo's David [Levoy 99]. After generating range images for all parts of an object's surface, these range images are combined into a single polygonal mesh with the help of surface reference points. The resulting 3D mesh completely describes an object to the extent that it is visible from the outside. Along with the range images photos of the surface are taken and mapped onto the combined mesh. To prepare the virtual model for interactive visualization light sources, material descriptions and a view point are now added to the mesh (see Fig. 6).

To finally view the virtual model and to allow interaction a software tool for 3D rendering and interactive visualization is needed. Possibilities for manipulation of virtual models by museum visitors can be to interactively change the view point in order to see an artwork from unusual directions, to change the lighting or to

compare it directly with other objects created by the same artist or in the same period of time. To make the 3D model accessible for Web users a conversion to the VRML format [VRML 97] is recommended. Users connecting to the museum's Web server can download the model and visualize it directly in their Web browser using a VRML plug-in.

Conclusion

Considering the whole digitization chain, scanning is the most important period. The error rate needs to be as small as possible to treat the original documents with the highest amount of care. We presented our methodology and quality criteria that proved successful to find a qualified scanning service for digitizing paper. We outlined especially those criteria, which we find helpful for other museums and libraries before starting their work. Although the investment in the highest image quality is necessary to build a future-proof digital collection, all the decisions described have to balance quality and cost.

The digitizing of Beethoven's heritage following the concepts described so far were planned to take place from the 1st of March 2000. During the starting phase however some problems caused a delay of several month. One of the major challenges for the digitizing service was to guarantee color fidelity in the whole digitization chain without loss of quality. Staff changes on the side of the scan service were necessary in order to engage a company that really was able to install ICC profiles [ICC 93] for each component of the digitization chain. In July 2000 new tests for different document types (images, prints, handwritings, letters) will proof true the color profiles, so that the production can be started as soon as possible. During the conference in October/November 2000, we will be able to present our approaches to the process.

However, the final proof of concepts will be achieved only while implementing and offering the full version of the 'Digital Beethoven House'.

References

[Bogen & Borowski 98] Bogen, M., Borowski, M: The Digital Beethoven House, AACE (Hrsg.): Proceedings of WebNet 98 - World Conference on the WWW, Internet & Intranet, 1998

[Bogen & Lahme & Lechner 99] Bogen, M., Lahme, S., Lechner, U.: The Digital Beethoven House: Conception and Demonstrator, INET'99: The 9th Annual Internet Society Conference, San Jose, California, USA, 22-25 June 1999, Proc., ISBN 1-891562-06-1, CD

[Chapman 99] Chapman, Stephen: Guidelines for Image Capture, Preservation Librarian for Digital Initiatives, Harvard University Library Preservation Center, February 1999, http://www.rlg.org/preserv/joint/chapman.html

[ICC 93] ICC Colormanagement, Home Page of the International Color Consortium (since 1993): http://www.color.org

[ISO9660 88] ISO 9660: 1988, Information processing -- Volume and file structure of CD-ROM for information interchange

[Kenney & Chapman 96] Kenney, A. R., Chapman, Stephen: Digital Imaging for libraries and archives, Department of Preservation and Conservation, Cornell University Library, June 1996

[Marshall] Marshall, Trevor G.: Color Tutorial, An easy to read explanation of Color Management, http://www.yarc.com/colortut.htm

[Mittler 98] Mittler, Elmar (Hrsg.): Retrospektive Digitalisierung von Bibliotheksbeständen, Berichte der von der Deutschen Forschungsgemeinschaft einberufenen Facharbeitsgruppen "Inhalt" und "Technik", Deutsches Bibliotheksinstitut, Berlin 1998

[Levoy 99] Levoy, M.: The Digital Michelangelo Project: Creating a 3D archive of his sculptures using laser scanning, Proceedings of Electronic Imaging and the Visual Arts '99, 1999

[MP3 93] MP3: ISO/IEC 11172-3:1993 Information technology -- Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s -- Part 3: Audio

[MPEG] MPEG: The MPEG Home Page: http://www.cselt.it/mpeg

[Noerr 98] Noerr, Peter: The Digital Library Tool Kit, Sun Microsystems, Inc., April 1998, p. 21.,

http://www.sun.com/products-n-solutions/edu/libraries/digitaltoolkit.html

[QUIC] QUIC: http://developer.apple.com/techpubs/quicktime/qtdevdocs/RM/pdfframe.htm

[REAL] REAL: RealAudio, RealVideo, http://www.realnetworks.com/

[RIFF] RIFF: Resource Interchange File Formats WAV, AVI:

http://www.ora.com/centers/gff/formats/micriff/index.htm#MICRIFF-DMYID.4

[Rigby 96] Rigby, Martin: What is FlashPix?, 1996, http://www.flashpix.com/WhatIsfpx.html

[TIFF 92] TIFF: Adobe Developers Association, TIFF (TM) Revision 6.0 - Final, June 3, 1992

[VRML 97] VRML: International Standard ISO/IEC 14772-1:1997, The Virtual Reality Modeling Language, http://www.vrml.org/

[WAVE] WAVE: The basics of wavelet: http://www.wavelet.org/